TOWARDS A SHARED LARGE-AREA MIXED REALITY SYSTEM

Naimul Mefraz Khan⋆ Xiaoming Nan⋆ Nan Dong⋆ Yifeng He⋆
Matthew Kyan† Jennifer James◦ Ling Guan⋆ Charles Davis◦

⋆ Department of Electrical and Computer Engineering, Ryerson University.
†Department of Electrical Engineering and Computer Science, York University.
◦ RTA School of Media, Ryerson University.

ABSTRACT

In this paper we present a large-area interactive mixed reality system where multiple users can experience an event simultaneously. Through the combination of a number of innovative methods, the system can tackle common problems that are inherent in most existing mixed reality solutions, such as robustness against lighting conditions, static occlusion, illumination correction, registration and tracking etc. Most importantly, with our proposed *experience server*, a shared event among multiple users is seamless. The experience server tracks every user's position and experience state and presents a unique viewpoint of the event to multiple users simultaneously. The effectiveness of the system is demonstrated through an example application at a heritage site, where we perform user testing through multiple focus groups.

Index Terms— Mixed Reality, Augmented Reality

1. INTRODUCTION

Virtual and augmented reality belong to the most exciting emerging technologies in recent years, with commercial market projected at 5 billion by 2016 [1].

Mixed reality (MR) is a combination of augmented and virtual reality [1]. We call our proposed framework a mixed reality system because on top of augmenting the virtual world onto the real world, our system changes the perception of reality by enabling a higher level of immersion through methods for occlusion handing, illumination correction and shared experience. Throughout the rest of the the paper, the discussion of the technologies are equally applicable to AR.

The advancement of mobile computing devices has opened up new opportunities and challenges in the world of MR. The powerful mobile devices enable the user to freely roam around in the world and experience MR events. However, providing the user a free reign is a technically challenging task because of the requirement of accurate alignment and registration between the virtual and the real world. Most popular commercial vision-based MR solutions (e.g. Vuforia ™[2]) require pre-defined markers and good lighting for accurate registration. These conditions not only restrict the user movement, but also becomes a hindrance in designing a good cinematic MR experience.

Another aspect of MR that is mostly overlooked is shared experience. Existing MR systems mostly focus on the registration and tracking aspect of MR for a single user. But an MR system that is aware of all the users experiencing an event opens up broader possibilities.

In this paper, we propose a mixed-reality system that tackles the aforementioned limitations. More specifically, the presented system has the following features:

- **Large Area:** Unlike most common MR solutions, our system does not restrict the user with pre-defined markers. As a statement to how large the area can be, our demo application was designed for a large room of size 10X10 meters.
- **Shared:** Our *Experience Server* enables seamless sharing of an event among the users simultaneously. This helps designing multitudes of different MR events which will be difficult to do under typical MR solutions focusing on a single-user experience.
- **Robust:** The tracking technology we use is not vision-based. Hence, we don't require good lighting or texture to design an MR experience. This robustness enables the creation of a cinematic experience with low lighting if required. In our demo application, the lighting was intentionally weaker than usual to create a dramatic ambiance.
- **Modular:** The design of the MR system is completely modular. The experience server has different modules (explained in detail later) for tracking and event sharing. Each of these modules can be used independently of the other. As a result, our system can be easily extended and modified to suit the need of the application.

We also provide an intelligent solution for dealing with static occlusion (Section 3.5).

E-mail address of first author: n77khan@ee.ryerson.ca

2. RELATED WORKS

Most of the existing literature focuses on three aspects of MR: devices, display and, most importantly, tracking/registration [3]. Our MR system focuses on the usability of MR by addressing other issues such as the area and shared nature. To make our system easily extendable, we have designed it for traditional mobile devices, billions of which are sold every year. By providing it as a common platform for mobile devices, we eliminate the need of specialized displays. Of course, due to the modular nature of the system, it can be used with specialized displays such as wearable computers or stereoscopic displays. But since that is beyond the scope of this paper, we only provide a brief overview of existing tracking technologies and the reasoning behind choosing the particular technology we used in our system.

Tracking/registration is the process of aligning virtual and real worlds together so that virtual objects appear to be part of the real world. The worlds are aligned by estimating the pose of the device. Two major types of tracking technologies currently exist, sensor-based and vision-based.

Sensor-based solutions use internal or external sensors to estimate the pose. Since most of the recent mobile devices are equipped with inertial sensors (e.g. accelerometer, gyroscope), they are a popular method of tracking for mixed reality [4]. However, these inertial sensors are prone to errors such as drift, lag and, hence, not suitable for a good MR experience [5]. To alleviate these issues, hybrid methods are often used [5], which fuses the inertial sensor information with others(e.g. visual) to provide better tracking.

Electromagnetic trackers use time of flight, received signal strength, phase difference etc. from electromagnetic signals to obtain pose information. Wifi/bluetooth are popular choices for electromagnetic signals [6], as they are commonly available. However, these signals are prone to ferromagnetic interference and are usually limited in range.

Another alternative is to use dedicated tracking systems for accurate pose information. For example, optical tracking systems such as OptiTrack ™[7] uses multiple IR cameras to localize passive markers that are placed on the devices. Although they are accurate to a few centimeters, they require the device to be visible by a minimum number of cameras and, hence, can result in putting a limit to the users' movement patterns. An alternative to optical systems are dedicated electromagnetic systems [8]. The same principal of tracking passive sensors is used, but instead of optical the emitted signal is high-frequency electromagnetic. This kind of tracking provides highly accurate pose information without putting a requirement on the visibility of the device. However, these dedicated solutions require elaborate hardware setup.

Vision-based solutions use feature correspondences from images extracted from the device camera. They can use predefined markers [9, 10] or natural features extracted during run-time [11, 12]. Since vision-based solutions only require a camera, they are highly portable.

The two major issues with vision-based solutions are computational processing time and dependence on environmental features [3]. While algorithmic improvements can alleviate the issues with computational time, the dependence on environmental features will never completely go away, as it is inherent to any vision-based tracking system. More specifically, for any vision-based tracking technology the environment will have to have enough textural information and reasonably illuminated. This limits the usability aspect of MR applications. The users can not freely roam around since loss of textural information will result in partial/complete loss of tracking. Even partial loss of tracking will hamper the user experience. Hence, the experiences are mostly designed in such a way that limits the user looking at a particular target area [10, 12]. In other words, they are not appropriate for a large area.

We wanted the users of our MR system to be able to physically walk around a space and enjoy different aspects of a large-area MR experience. We also did not want to put any restriction on the lighting and texture requirements, as that can deteriorate the theatrical value of an event. For these reasons, we used a dedicated electromagnetic system for tracking [13]. Although they are not a popular solution for MR due to the requirement of elaborate hardware setup, as we will see, using a robust tracking system opens up new possibilities in the experience design which otherwise will not be possible.

In our experiments, we found that solely using an electromagnetic system is not practical, and even a dedicated system is prone to some jitter. Hence, our final tracking system (explained in detail in Section 3.2) consisted of a hybrid tracker which relies on both inertial sensors and electromagnetic tracker to provide a natural experience.

3. THE MR SYSTEM

3.1. Experience Server

Figure 1 shows the overall architecture of our system. As we can see, the passive sensors are attached to the client devices. The hardware tracker receives pose information from the sensors in real time.

The heart of the system is the *Experience Server*. The experience server has two modules. The pose tracker does necessary calculations to convert the raw data received from the hardware tracker into meaningful pose information. This includes converting the coordinate systems between the tracking data and the virtual world into a common MR world coordinate system, synchronizing the data among multiple clients etc.

The state tracker enables the unique sharing aspect of our system. Essentially, the state tracker monitors the experience state of each client. Everything that happens in the virtual world is controlled by the state tracker. The state tracker also

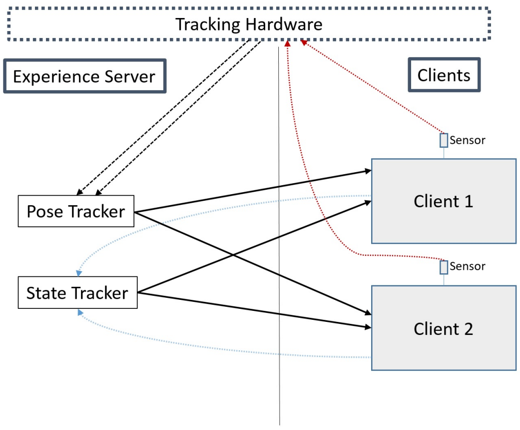


Fig. 1. Overall System Architcture.

has real-time pose information for every client obtained from the pose tracker. As a result, the state tracker is aware of where the the clients are positioned in the MR world. Hence, the flow of the event can be controlled and synchronized among multiple users easily and efficiently. The example application in Section 4 will provide a clear example of how the state tracker enables a shared MR experience.

3.2. Hybrid Tracking

As discussed before, using the electromagnetic tracking system directly can result in unwarranted noise and jitter. As a result, we used a hybrid tracking method, where information from both the inertial sensors and the hardware tracking system were combined.

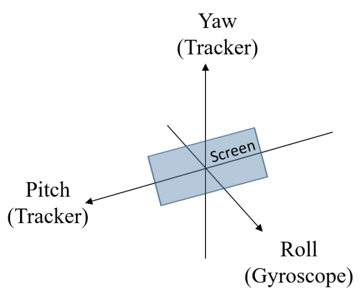


Fig. 2. Hybrid tracking for rotational adjustment.

The inertial sensors does not provide any advantage in terms of positional information, since the hardware tracking system already provides excellent accuracy. We employed a low-pass filter on the hardware tracking data to reduce some jitter.

There are three components to the rotation, roll, pitch, and yaw (Figure 2). During the experiments, we found that the users are extremely sensitive to any jitter in roll. This is expected, since jitter in roll will result in misalignment between the horizontal axes (the grounds) of the real and the virtual world, which is easily noticeable. Hence, we obtained the roll from the gyroscope of the device. The roll and pitch were obtained from the hardware tracker (after low-pass filtering). Gyroscope data was also used to initially align the virtual and real world axes systems.

3.3. Device Camera Calibration

To provide a properly scaled alignment between the virtual and the real world, the field of view of the virtual camera and the real camera need to be matched. Field Of View (FOV) describes the angular extent of a given scene that is currently viewable by the camera [14]. We calculated the FOV of device camera and set the virtual camera FOV to the same value. Without matching these two values accurately, a viewer will have a drifting effect while moving around in the real space.

Most devices do not list the field-of-view as an official spec, especially in video mode. Hence, we estimated it through experimentation. By using a measured checkerboard pattern, the focal length of the device camera can be calculated using a a four step camera calibration procedure [15]. Once we can find the focal length, the FOV can be calculated using the following equation [14]:

$$\mathbf{V} = 2 * arctan\left(\frac{S}{2 * \mathbf{F}}\right), \tag{1}$$

where, $\mathbf{F}$ is the focal length calculated from camera calibration. S is the aspect ratio of the device screen.

3.4. Illumination Correction

The illumination condition in the real world can be very different from that of the virtual world, which can deteriorate the quality of an MR experience. In order to address this issue, we apply a contrast stretching method to process the real images in real time. Specifically, we change the contrast in an image by stretching/shrinking the range of intensity values it contains to match the desired range of values in the virtual world. In the first step, we need to determine the intensity range of virtual world over which the image pixel values will be extended. Multiple light sources are integrated in a virtual scene, including point light, directional light, and ambient light. We examine the virtual objects under the synthetic lighting and find the upper bound value V^c_{max} and the lower bound value V^c_{min} for each channel c ($c \in \{red, green, blue\}$). When capturing the real word scene through the camera feed, the upper bound R^c_{max} and the lower bound R^c_{min} of the intesity range is determined ($c \in \{red, green, blue\}$). If $R^c_{max} - R^c_{min} < V^c_{max} - V^c_{min}$, the intensity values in real images will be stretched to fit the virtual range, while the real intensity range will be compressed if $R^c_{max} - R^c_{min} > V^c_{max} - V^c_{min}$. Each pixel P^c_{in} in the original image will be mapped to output value P^c_{out} using the function:

$$P^c_{out} = (P^c_{in} - R^c_{min}) \left(\frac{V^c_{max} - V^c_{min}}{R^c_{max} - R^c_{min}} \right) + V^c_{min}, \qquad (2)$$

where $c \in \{red, green, blue\}$.

3.5. Static Occlusion

Occlusion refers to the blocking of virtual objects by objects in the real world. Without occlusion handling, virtual objects sit on top of real objects, like projected images. Inaccurate occlusions can greatly deteriorate the experience.

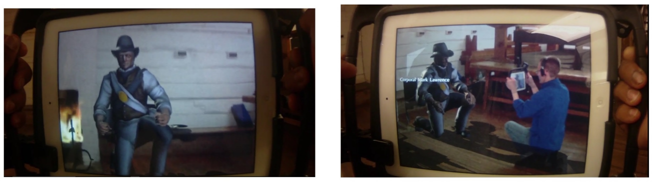
Real-time occlusion estimation for dynamic environments is a challenging research topic, and most existing solutions put some form of restriction on the environment or the mobility of the user [16]. In the proposed system, we use an intelligent technique to mask real world objects efficiently. To calculate the occlusion factor resulting from a real-world object, a 3D model for the object is created in the virtual world. The stencil buffer was then used to create a mask using the 3D model. The camera feed passed through the mask to create the illusion that the virtual objects are occluding the real world.

For the prototype system, we assumed that the real environment is static, which will not be possible in an uncontrolled setting. In future, we plan to investigate incorporating dynamic segmentation methods into the system.

4. EXAMPLE APPLICATION

The proposed MR system was tested with deployment at a heritage site (undisclosed to maintain anonymity). The MR event took place in a large room of size 10X10 meters. The users were free to roam around anywhere in the room. Different historical characters were placed across the room. The event itself was a dramatic recreation of what happened in that room a few hundred years ago. Five users could experience the event simultaneously. A number of the virtual characters were interactive, where they would approach the users and have a conversation with them when they were in a proximity. Basic interaction elements such as head turning to follow the user, voice recognition etc. were implemented with the help of off-the-shelf software. The experience was created with Unity3D ™game engine [17]. For electromagnetic tracking, the Polhemus G4 ™[13] system was used. The

Fig. 3. Screenshot of the experience server for the example application. The tiny rectangles represents current locations of the users.



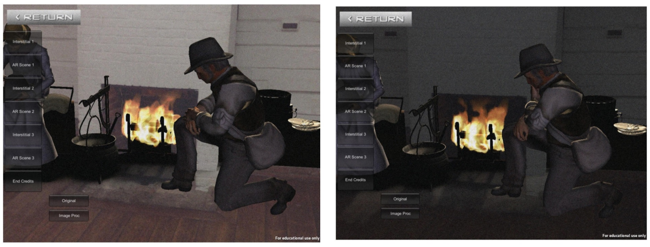
(a) (b)

Fig. 4. Example of a shared experience. (a) First person view of talking to a character. (b) Third-person view of the same event synchronously.

users were provided with sensor-mounted iPad 4 devices and headphones.

Although the RMS tracking accuracy reported by Polhemus is 2 mm [13], in practice we found the accuracy varies between 6 mm - 10 mm. This is because in a practical environment, the electromagnetic system will have interference from ferromagnetic materials. The room that the experience was designed for had water pipes running across the ceiling where the tracking system was mounted to. However, even 6-10 mm is a very high level of accuracy for MR, and the user feedback regarding tracking was satisfactory.

Figure 3 shows a screenshot of the experience server. The tiny rectangles represent the current position and orientation of each user's device in the MR world. The virtual masks created for static occlusion (e.g. the staircase) can also be seen in the figure. The experience server was deployed in the control room where the administrator could visualize the entire state of the event quickly.



(a) (b)

Fig. 5. Illustration of contrast stretching in the proposed system. (a) Without contrast stretching. (b) With contrast stretching.

Figure 4 shows an example of a shared experience. Figure 4-(a) shows the first-person view of a user talking to a virtual character. 4-(b) shows a third person witnessing the conversation. 4-(b) is an excellent example of what our robust experience server can enable. With most popular MR solutions, it will not be possible to create such an immersive experience as seamlessly as it is shown here. The user feedback for this form of shared events was overwhelmingly positive.

Figure 5 illustrates the result of contrast stretching in the proposed mixed reality system. The result without contrast stretching is displayed in Figure 5-(a), in which the real world background is much brighter compared to the virtual character. 5-(b) shows the mixed scene with contrast stretching. We can see that the intensity range in the real world is compressed to match the virtual world and thus the virtual character is conspicuous in the dark ambiance.

(a) (b)

Fig. 6. Handling of static occlusion in the proposed system. (a) Without static occlusion. (b) With static occlusion.

Figure 6 shows the result of our static occlusion handling. We can see that the virtual drums are always drawn on top of the real staircase without occlusion handling (Figure 6 -(a)). Our static occlusion handling process does reasonably well to place the virtual drums in the right place in the MR world. In future, we will investigate incorporating automatic segmentation methods to provide a more robust occlusion handling solution.

4.1. User Experience (UX) Testing

The most unique aspect of the system is the ability to create shared experiences. Hence, during UX testing, a group of users ran through the experience together rather than individually. The methodology to evaluate their experience was focus groups. After the experience was finished, the group of users participated in a discussion session where a set of questionnaire focusing on different aspects of the system was handed out. The uniqueness of the shared aspect of the experience made it difficult to devise quantitative parameters for UX testing. Hence, the questions were mostly qualitative. In future we plan to do further controlled and quantitative user testing. Some example questions were:

- How did having four other people in the room influence your interaction with the story going on around you (*shared experience*)?
- Which specific points of the experience were the best and the worst examples of the registration and blending between the virtual and the real world (*illumination/occlusion*)?

We tested the experience with with six focus groups, each group consisting of 4-5 users. The groups were carefully chosen to cover varying demographics. For example, The first focus group consisted of participants between the ages of 20 and 30; 4 female participants and 1 male participant. The second focus group consisted of children between the ages of 8 and 12; 2 female children and 2 male children. The other focus groups consisted of young adults, middle-aged participants, all male users etc.

In general, the feedback for the experience was very positive. For many of the users, this was the first mixed reality experience and they were thrilled by it. Even for users who have experienced some form of mixed/augmented reality, the large area and the opportunity to roam around freely greatly enhanced their experience.

All of the users were positive about the shared nature of the experience. The users especially mentioned that experiencing it in a group enhanced the interactivity. The users were especially impressed by the scenario presented in Figure 4, where one user interacts with the virtual character while the others witness it through their respective devices. One negative point about the shared experience was that some of the users were not aware of the event taking place since they were busy in another corner of the room. However, this is a limitation of the example application narrative rather than the technology itself. A carefully designed narrative will be able to draw all the users to the most exciting part of the experience.

Most of the users noticed the application of our illumination correction method and agreed that it had a positive effect on the overall experience.

In terms of occlusion, users were impressed by the immersive nature of our static occlusion procedure. However, in some cases the virtual mask did not properly align with the real world. We are investigating whether it is because of noise in tracking, or improper scaling of the mask.

The biggest drawback of the shared nature of the experience is the lack of dynamic occlusion. The users complained about other real-world users getting in the way of the mixed reality world. As stated before, dynamic occlusion is a challenging problem to solve and beyond the scope of this paper. In future, we will add a separate module to the system for dynamic occlusion handling.

5. CONCLUSION

In this paper, we presented a large-area MR system. Our proposed MR system addresses numerous issues that can hinder an MR experience, such as robust tracking, illumination correction, occlusion handling etc. The most unique aspect of the system is the seamless sharing of events among multiple users simultaneously. Our experience server can track the location and state of each user in the system and synchronize their experience accordingly. Our proposed system is also modular, so that new experiences can be created quickly and efficiently. We showed the effectiveness of our proposed system through an example application, where multiple users experienced a virtual recreation of history at a prominent heritage site. The feedback from the users were generally positive with particular appreciation of the shared nature of the experience.

In future, we will work on improving different aspects of the system further, especially occlusion handling.

6. REFERENCES

[1] M. Billinghurst, A.Clark, and G. lee, "A survey of augmented reality," *Foundations and Trends in Human-Computer Interaction*, vol. 8, no. 1, pp. 1–202, 2015.

[2] PTC Inc., "Vuforia," `https://developer.vuforia.com/`, 2015.

[3] Z. Huang, P. Hui, C. Peylo, and D. Chatzopoulos, "Mobile augmented reality survey: a bottom-up approach," *CoRR*, vol. abs/1309.4413, 2013.

[4] S. White, S. Feiner, and J. Kopylec, "Virtual vouchers: Prototyping a mobile augmented reality user interface for botanical species identification," in *3D User Interfaces*, 2006, pp. 119–126.

[5] X. Yang, X. Si, T. Xue, L. Zhang, and K.-T. Cheng, "Vision-inertial hybrid tracking for robust and efficient augmented reality on smartphones," in *ACM Multimedia*, 2015, pp. 1039–1042.

[6] R. Wang, F. Zhao, and H. Y. Luo, "Fusion of wi-fi and bluetooth for indoor localization," in *LMBS*, 2011, pp. 63–66.

[7] NaturalPoints Inc., "OptiTrack," `http://www.optitrack.com/`, 2015.

[8] J. Huang, T. Mori, K. Takashima, S. hashi, and Y. Kitamura, "IM6D: magnetic tracking system with 6-dof passive markers for dexterous 3D interaction and motion," *ACM Transactions on Graphics*, vol. 34, no. 6, 2015.

[9] N. Hagbi, O. Bergig, , and J. E. Sana, "Shape recognition and pose estimation for mobile augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 65–71.

[10] C.-Y. Chen, B.R. Chang, and P.-S. Huang, "Multimedia augmented reality information system for museum guidance," *Personal and Ubiquitous Computing*, vol. 18, pp. 315–322, 2014.

[11] J. Lei, Z. Wang, Y. Wu, and L. Fan, "Efficient pose tracking on mobile phones with 3d points grouping," in *IEEE International Conference on Multimedia and Expo*, 2014, pp. 1–6.

[12] D. Molyneaux and S. BenHimane, "It's a Pirates Life: AR Game," in *IEEE International Symposium on Mixed and Augmented Reality*, 2014, pp. 361–362.

[13] Polhemus, "Polhemus G4," `http://polhemus.com/motion-tracking/all-trackers/g4`, 2015.

[14] D.A. Forsyth and J. Ponce, Eds., *Computer Vision: A Modern Approach*, Prentice Hall, New York, NY, USA, 2002.

[15] J. Heikkila and O. Silven, "A four step camera calibration procedure with implicit image correction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106–1112.

[16] G. Xing, Y. Liu, X. Qin, and Q. Peng, "On-line illumination estimation of outdoor scenes based on area selection for augmented reality," in *International Conference on Computer-Aided Design and Computer Graphics (CAD/Graphics)*, 2011, pp. 439–442.

[17] Unity Technologies, "Unity3D," `http://www.unity3d.com/`, 2015.